

Impact case study (REF3)

Institution: University of Manchester		
Unit of Assessment: 11 (Computer Science and Informatics)		
Title of case study: Context aware text mining for the pharmaceutical sector		
Period when the underpinning research was undertaken: 2010 – 2015		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s): Goran Nenadic	Role(s) (e.g. job title): Professor (2017 – present), Senior Lecturer (2004 – 2017)	Period(s) employed: 2004 – present
David Robertson	Professor	2002 – 2016
Period when the claimed impact occurred: 2014 – July 2020		
Is this case study continued from a case study submitted in 2014? N		
1. Summary of the impact		
<p>The University of Manchester has pioneered research into context-aware text mining: methodologies to increase the utility and reliability of scientific claims automatically extracted from large-scale document repositories, such as scientific bibliographic databases.</p> <p>The research underpinned the formation of a new business, Biorelate Ltd., and their main product Galactic-AI™. Since 2014, Biorelate Ltd have increased the efficacy and economic viability of several companies, both small and large, in the pharmaceutical industry (e.g. AstraZeneca).</p> <p>[Text removed for publication]</p>		
2. Underpinning research		
<p>The worldwide output of scientific papers approximately doubles every nine years. As a result, new drugs, for example, are often developed without a full view of the literature. In response to this, the text mining community developed methods to automatically extract textual claims of interactions between biological entities, such as drugs, genes, and proteins. The generic word for such an interaction is an `event`, e.g. "<i>gene X is regulated by gene Y</i>". Knowledge of such events aids in the development of new drugs and treatments, providing reassurance of new hypotheses, and generation of new ones. However, these text mining methods were limited in their ability to contextualise the event, such as whether it occurs only in a particular species, or particular organ, or under specific conditions.</p> <p>In 2010, research led by Nenadic [1] started to address this problem, developing methods to automatically extract certain contextual information about the event, such as</p> <ul style="list-style-type: none"> • in which biological species the event occurred, or, • which anatomical location the claim applies to. <p>A key challenge here is that contextual information could be spread throughout a paper, and thus requires long-distance linking to identify chains of events. The automatic extraction of such information involved development of novel combinations of deep linguistic parsing and event extraction methods [1]. These methods (i) introduced a new approach to event integration across the document by tracing and linking equivalent events, and (ii) reduced false positives through post-processing based on an analysis of the outputs - e.g. identifying</p>		

Impact case study (REF3)

that circularly nested event chains were likely false positives. This resulted in the text-mining system BioContext [1] (see Figure 1).

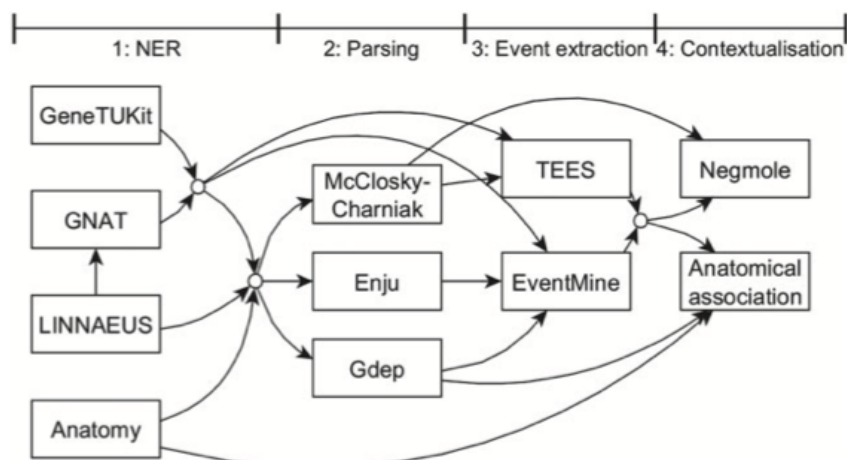


Figure 1. The system architecture of the text-mining tool BioContext

Between 2012 and 2014, further research [2,3] added bibliographic and logical contexts to domain-specific contextual information, including

- **where** the event was claimed (e.g. publication venue), and by whom;
- in **what** part of a document it was claimed (e.g. abstract or conclusions);
- whether or not (and how) others may have **refuted** it at a later point *in time*.

An important point to note is the final one on refutation that illustrates that a claimed event may or may not be **reproducible research** in the long run.

The addition of context enables the *prioritisation* of the many millions of extracted event claims, giving a rank order in terms of the confidence that can be attributed to the claim, as well as using the context to distinguish cause vs. correlation in the events. This approach was first applied to support development of the human immunodeficiency virus type 1 (HIV-1) human protein interaction database [2]. The approach identified and prioritised a number of missing events that should be added to the database (a 6-fold increase in the number of unique interactions). Results demonstrated that most false positives were not caused by falsely reported information, but due to an *incomplete* event chain. The addition of context permitted reconstruction of *complete* event chains, thus reducing false positives [2].

In [3], the methodology was evolved, contributing (i) a new method for document relevance-scoring to individual concepts (a weighted sum of task-relevant and task-specific terms) and (ii) new methods for determining the relevance of an event chain to disease terms (based on the ordering of, and character distance between, mentions within the document). This was demonstrated on a text corpus on chronic pain, a complex phenomenon with a large societal burden and unmet medical need. A total of 762,692 research articles were processed, resulting in 356,499 ranked unique contextualised events [3]. In [4], the methodology was applied (in collaboration with Pfizer) in the Chronic Pain community, demonstrating the applicability of this research in the biomedical community.

Impact case study (REF3)

3. References to the research

Part of the research [2,3,4] was funded through a BBSRC/Case partnership with Pfizer in 2010 (GBP75,000). [1-3] are published in top journals in the field (both Q1 journals). All citations are from Google Scholar (November 2020).

- [1] Gerner M, Sarafraz F, Bergman CM, **Nenadic G** (2012) BioContext: an integrated text mining system for large-scale extraction and contextualization of biomolecular events, *Bioinformatics*, **28** (16) [DOI: 10.1093/bioinformatics/bts332](https://doi.org/10.1093/bioinformatics/bts332) (63 Citations)
- [2] Jamieson, DG, Gerner, M, Sarafraz, F, **Nenadic, G & Robertson, DL** (2012) Towards semi-automated curation: Using text mining to recreate the HIV-1, human protein interaction database, *Database*, vol. 2012, bas023. [DOI: 10.1093/database/bas023](https://doi.org/10.1093/database/bas023) (25 Citations)
- [3] Jamieson, DG, Roberts, PM, **Robertson, DL**, Sidders, B & **Nenadic, G** (2013) Cataloguing the biomedical world of pain through semi-automated curation of molecular interactions, *Database*, vol. 2013, [DOI:10.1093/database/bat033](https://doi.org/10.1093/database/bat033) (15 Citations)
- [4] Jamieson, DG, Moss, A, Kennedy, M, Jones, S, **Nenadic, G, Robertson, DL & Sidders, B** (2014), The pain interactome: Connecting pain-specific protein interactions, *Pain*. **155** (11) [DOI: 10.1016/j.pain.2014.06.020](https://doi.org/10.1016/j.pain.2014.06.020) (25 Citations)

4. Details of the impact**Pathway to impact**

In 2014, drawing upon the underpinning research on contextualised text mining [1-4], Biorelate Ltd was founded by Daniel Jamieson, a University of Manchester (UoM) PhD alumnus, supervised by Nenadic and Robertson. Biorelate used the processes and techniques developed in [1-4] to develop an innovative new product, Galactic-AI™, which Biorelate confirm is: *“a cloud-based cognitive computing platform that automatically curates text articles using a pipeline of deep learning and [natural language processing] NLP software services, all of which would not have been possible without the original research conducted at Manchester”* [A].

At the company's formation, Nenadic was appointed as scientific advisor to the company, and retains this role to date.

Direct economic impact: foundation of a new business

In 2014, Biorelate Ltd was formed as a new technology-led business, with the aim to bring its new product to the pharmaceutical sector. Its main product, Galactic-AI™, uses the technology identified in [1-4] to enable the rapid and reliable curation of event claims extracted from large bodies of literature (millions of articles, hundreds of millions of relationships), and generation of novel scientific hypotheses, at a scale that was previously economically unviable. The platform is currently capable of processing 50,000,000 articles in 12 hours [A].

Biorelate works on a contractual basis; each contract is tailored to the specific scientific goals of the individual client.

[Text removed for publication]

Catapult Ventures confirm that Biorelate Ltd have *“developed the technology, building on revenue generating projects and validated the business model... [Catapult] are excited to provide further investment and to support the company's ambitions to become a leader in the curation of Biomedical knowledge”* [C].

Impact case study (REF3)

[Text removed for publication]

In May 2020, Biorelate provided free registration to Galactic Web (the online web-interface of the Galactic-AI product) to support researchers impacted by the COVID-19 crisis and lockdowns. This was to ensure researchers who were restricted in access to labs could still undertake important industrial research. Over the first COVID-19 lockdown period, this saw a 10-fold increase in the number of user sign-ups across the life sciences [A].

Indirect economic impacts: contributing to innovation through the delivery of new products and services in the pharmaceutical sector

Through underpinning a new product that has improved the efficiency of pharmaceutical hypothesis generation, the UoM research [1-4] has achieved significant impact within pharmaceutical companies, which have used the technology to enhance and diversify their business activities. As a result, benefits to drug treatment and diagnosis companies include:

(i) e-Therapeutics: developing treatment for fibrosis

e-Therapeutics – a UK based specialist company that undertakes network-driven drug discovery – first partnered with Biorelate in 2016 and uses Galactic-AI™ for two distinct activities within their own platform and drug discovery processes. At specific stages of a discovery project, e-Therapeutics conducts many *ad-hoc* reviews of the literature. The Chief Technical Officer of e-Therapeutics has confirmed that *“the technology approximately doubles our team's efficiency in performing this task, saving up to 1,000 person-hours a year. In financial terms I would estimate this as GBP50,000 per year in staff time. The technology is therefore a significant asset to our business and a key part of our plans to expand and scale.”* [D].

e-Therapeutics uses Galactic-AI™ as part of their fibrotic disease discovery programme, and confirms it thus enabled them *“to develop new hypotheses for the treatment of fibrosis, which we are confident we would not have been able to develop in an economically viable time-scale, if it were not for being able to use Galactic-AI”* and adding that it *“has significantly influenced our business activities, providing new directions for several highly skilled staff.”* [D].

e-Therapeutics currently employ 10 staff to work directly with the Biorelate technology. The significance of this partnership is evident in the e-Therapeutics share price increasing by 3.7% immediately following the announcement of their partnership with Biorelate on 15 January 2018 [D]. With 421,000,000 shares in issue, this equates to an increase in shareholder wealth of GBP1,515,600 [E].

(ii) Apconix: supporting Target Safety Assessments for drug discovery

Apconix, global experts in non-clinical drug safety, provide target safety assessments (TSAs) to drug discovery companies to reduce the risk of failure in drug discovery and development. A TSA is a concise review of data presented as actionable recommendations for the safety evaluation strategy and is critical to the success of early discovery programmes. TSAs are fundamental for decision making, understanding potential risks and prioritising further drug research activities.

Apconix approached Biorelate in 2017, and have used Galactic-AI™ to create and market their own new product – Acuity – an AI-enabled TSA. Galactic-AI™ is used to auto-curate and organise the knowledge that is required to conduct a TSA. The Apconix co-founder has confirmed *“without our partnership with Biorelate, and their product Galactic-AI, we would not have been able to bring our product, Acuity, to market”* [F]. Further, Apconix have experienced a 40% increase in the number of TSA contracts they have undertaken, attributed to the Biorelate collaboration: *“As a direct result of our collaboration with Biorelate, we have 80 contracts for TSAs in 2020 compared with a total of 48 in 2019”* [F].

Impact case study (REF3)

(iii) [Text removed for publication]

(iv) [Text removed for publication]

5. Sources to corroborate the impact

- [A] Letter of support from the Chairman of Biorelate, November 2020.
- [B] Press Release confirming Innovate funding to Biorelate UK, February 2019
- [C] Manchester Evening News confirming funding to Biorelate UK, September 2018
- [D] Letter of Support from Chief Technical Officer of e-Therapeutics, October 2020
- [E] Calculation of share price increase - pdf of sources and calculation undertaken
- [F] Letter of Support from Co-director of Apconix, October 2020
- [G] Letter of Support from Associate Director (Oncology Bioinformatics, Translational Medicine) of AstraZeneca, March 2020
- [H] Sidders, B., Karlsson, A., Kitching, L., Torella, R., Karila, P and Phelan, A. (2018) Network-Based Drug Discovery: Coupling Network Pharmacology with Phenotypic Screening for Neuronal Excitability, *Journal of Molecular Biology*, 430 (18), 3005-3015, DOI: [10.1016/j.jmb.2018.07.016](https://doi.org/10.1016/j.jmb.2018.07.016)
- [I] Sidders B, Zhang P, Goodwin K, O'Connor G, Russell DL, Borodovsky A et al., (2020) Adenosine signalling is prognostic for cancer outcome and has predictive utility for immunotherapeutic response, *Clinical Cancer Research*. 26 (9), DOI: [10.1158/1078-0432.CCR-19-2183](https://doi.org/10.1158/1078-0432.CCR-19-2183)